

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
4 October 2001 (04.10.2001)

PCT

(10) International Publication Number  
WO 01/73128 A1

(51) International Patent Classification<sup>7</sup>: C12Q 1/68,  
C12P 19/34; C07H 21/04

(74) Agents: BUTLER, James, E. et al.; Senniger, Powers,  
Leavitt & Roedel; One Metropolitan Square, 16th Floor,  
St. Louis, MO 63102 (US).

(21) International Application Number: PCT/US01/09583

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 26 March 2001 (26.03.2001)

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

Published:

— with international search report

(26) Publication Language: English

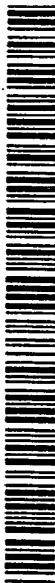
For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(30) Priority Data:  
60/191,737 24 March 2000 (24.03.2000) US

(71) Applicant (for all designated States except US): DZ-  
GENES, LLC [US/US]; 6420 Clayton Road, Mother  
Concordia Hall, Ground Floor, Richmond Heights, MO  
63117 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): MOSKOWITZ,  
David, W. [US/US]; 518 Bonhomme Woods Drive, St.  
Louis, MO 63132 (US).



WO 01/73128 A1

(54) Title: DIAGNOSTIC POLYMORPHISMS OF TGF- $\beta$ -RII PROMOTER

(57) Abstract: Disclosed are single nucleotide polymorphisms (SNPs) associated with hypertension and end stage renal disease due to hypertension. Also disclosed are methods for using SNPs to determine susceptibility to end stage renal disease and hypertension; nucleotide sequences containing SNPs; kits for determining the presence of SNPs; and methods of treatment or prophylaxis based on the presence of SNPs.

## DIAGNOSTIC POLYMORPHISMS OF TGF- $\beta$ -RII PROMOTER

### CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. provisional application serial No. 60/191,737, filed March 24, 2000, which is incorporated herein by reference in its 5 entirety.

### BACKGROUND

This invention relates to detection of individuals at risk for pathological conditions based on the presence of single nucleotide polymorphisms (SNPs).

During the course of evolution, spontaneous mutations appear in the genomes 10 of organisms. It has been estimated that variations in genomic DNA sequences are created continuously at a rate of about 100 new single base changes per individual (Kondrashow, *J. Theor. Biol.*, 175:583-594, 1995; Crow, *Exp. Clin. Immunogenet.*, 12:121-128, 1995). These changes, in the progenitor nucleotide sequences, may 15 confer an evolutionary advantage, in which case the frequency of the mutation will likely increase, an evolutionary disadvantage in which case the frequency of the mutation is likely to decrease, or the mutation will be neutral. In certain cases, the mutation may be lethal in which case the mutation is not passed on to the next generation and so is quickly eliminated from the population. In many cases, an equilibrium is established between the progenitor and mutant sequences so that both 20 are present in the population. The presence of both forms of the sequence results in genetic variation or polymorphism. Over time, a significant number of mutations can accumulate within a population such that considerable polymorphism can exist between individuals within the population.

Numerous types of polymorphism are known to exist. Polymorphisms can 25 be created when DNA sequences are either inserted or deleted from the genome, for example, by viral insertion. Another source of sequence variation can be caused by the presence of repeated sequences in the genome variously termed short tandem repeats (STR), variable number tandem repeats (VNTR), short sequence repeats (SSR) or microsatellites. These repeats can be dinucleotide, trinucleotide,

tetranucleotide or pentanucleotide repeats. Polymorphism results from variation in the number of repeated sequences found at a particular locus.

By far the most common source of variation in the genome are single nucleotide polymorphisms or SNPs. SNPs account for approximately 90% of 5 human DNA polymorphism (Collins et al., *Genome Res.*, 8:1229-1231, 1998). SNPs are single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in a population. Several definitions of SNPs exist in the literature (Brooks, *Gene*, 234:177-186, 1999). As used herein, the term "single nucleotide polymorphism" or "SNP" includes all single base variants and so 10 includes nucleotide insertions and deletions in addition to single nucleotide substitutions (e.g. A->G). Nucleotide substitutions are of two types. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine for a pyrimidine or vice versa.

15 The typical frequency at which SNPs are observed is about 1 per 1000 base pairs (Li and Sadler, *Genetics*, 129:513-523, 1991; Wang et al., *Science*, 280:1077-1082, 1998; Harding et al., *Am. J. Human Genet.*, 60:772-789, 1997; Taillon-Miller et al., *Genome Res.*, 8:748-754, 1998). The frequency of SNPs varies with the type and location of the change. In base substitutions, two-thirds of the substitutions 20 involve the C->T (G->A) type. This variation in frequency is thought to be related to 5-methylcytosine deamination reactions that occur frequently, particularly at CpG dinucleotides. In regard to location, SNPs occur at a much higher frequency in non-coding regions than they do in coding regions.

SNPs can be associated with disease conditions in humans or animals. The 25 association can be direct, as in the case of genetic diseases where the alteration in the genetic code caused by the SNP directly results in the disease condition. Examples of diseases in which single nucleotide polymorphisms result in disease conditions are sickle cell anemia and cystic fibrosis. The association can also be indirect, where the SNP does not directly cause the disease but alters the 30 physiological environment such that there is an increased likelihood that the patient will develop the disease. SNPs can also be associated with disease conditions, but play no direct or indirect role in causing the disease. In this case, the SNP is located

close to the defective gene, usually within 5 centimorgans, such that there is a strong association between the presence of the SNP and the disease state. Because of the high frequency of SNPs within the genome, there is a greater probability that a SNP will be linked to a genetic locus of interest than other types of genetic markers.

5 Disease associated SNPs can occur in coding and non-coding regions of the genome. When located in a coding region, the presence of the SNP can result in the production of a protein that is non-functional or has decreased function. More frequently, SNPs occur in non-coding regions. If the SNP occurs in a regulatory region, it may affect expression of the protein. For example, the presence of a SNP 10 in a promoter region may cause decreased expression of a protein. If the protein is involved in protecting the body against development of a pathological condition, this decreased expression can make the individual more susceptible to the condition.

Numerous methods exist for the detection of SNPs within a nucleotide sequence. A review of many of these methods can be found in Landegren et al., 15 *Genome Res.*, 8:769-776, 1998. SNPs can be detected by restriction fragment length polymorphism (RFLP) (U.S. Patent Nos. 5,324,631; 5,645,995). RFLP analysis of the SNPs, however, is limited to cases where the SNP either creates or destroys a restriction enzyme cleavage site. SNPs can also be detected by direct sequencing of the nucleotide sequence of interest. Numerous assays based on hybridization have 20 also been developed to detect SNPs. In addition, mismatch distinction by polymerases and ligases has also been used to detect SNPs.

There is growing recognition that SNPs can provide a powerful tool for the detection of individuals whose genetic make-up alters their susceptibility to certain diseases. There are four primary reasons why SNPs are especially suited for the 25 identification of genotypes which predispose an individual to develop a disease condition. First, SNPs are by far the most prevalent type of polymorphism present in the genome and so are likely to be present in or near any locus of interest. Second, SNPs located in genes can be expected to directly affect protein structure or expression levels and so may serve not only as markers but as candidates for gene 30 therapy treatments to cure or prevent a disease. Third, SNPs show greater genetic stability than repeated sequences and so are less likely to undergo changes which would complicate diagnosis. Fourth, the increasing efficiency of methods of

detection of SNPs make them especially suitable for high throughput typing systems necessary to screen large populations.

One disease for which the discovery of markers to detect increased genetic susceptibility is critically needed is end-stage renal disease. End-stage renal disease 5 (ESRD) is defined as the condition when life becomes impossible without replacement of renal functions either by kidney dialysis or kidney transplantation. Hypertension (HTN) and non-insulin dependent diabetes (NIDDM) are the leading causes of end-stage renal disease (ESRD) nationally (United States Renal Data System, Table IV-3, p. 49, 1994). There is currently an epidemic of ESRD, due 10 mainly to the aging of the American population. The ESRD epidemic is of special concern among African Americans where the incidence of ESRD is four- to six-fold higher than for Caucasians (Brancati et al., *J. Am. Med. Assoc.*, 268:3079-3084, 1992), but where treatment of hypertension, a causative factor in ESRD, is less effective (Walker et al., *J. Am. Med. Assoc.*, 268:3085-3091, 1992).

15 There are currently 200,000 patients with ESRD receiving renal replacement therapy (dialysis or renal transplantation), with an annual cost of \$13 billion. These numbers will certainly increase as the population of the nation continues to age. Since 1980, when complete data became available for the first time, most new cases 20 of ESRD have been ascribed to NIDDM or hypertension. The incidence of ESRD due to NIDDM or hypertension is still increasing, suggesting that the U.S. is in the early phase of an epidemic of ESRD. Preventing ESRD would save at least \$30,000 per patient, per year in dialysis costs alone, as well as enhance the patient's quality of life and ability to work. It is clearly the ideal method of cost-containment for renal disease. Without effective prevention of ESRD, the nation will instead be 25 forced to adopt less humane methods of cost-containment, such as denial of access (gate-keeping), or rely upon unrealistic expectations about patient reimbursement rates, etc.

Transforming growth factor beta (TGF- $\beta$ ) is a multifunctional polypeptide growth factor implicated in a variety of renal diseases. Almost every cell in the 30 body has been shown to make some form of TGF- $\beta$ , and almost every cell has receptors for TGF- $\beta$ , the context of which determines their functionality. The transforming growth factor- $\beta$  system is also a likely mediator of renal apoptosis.

TGF- $\beta$  is intimately connected with glomerular sclerosis, mesangial matrix expansion, and tubulointerstitial fibrosis in experimental rodent models and human glomerulonephritis (Border et al., *Kidney Int.*, 47 (Suppl. 49):S-59-S-61, 1995). Of the three isoforms available, TGF- $\beta$ 1 has been implicated most consistently in pathologic fibrosis (Khalil et al., *Am. J. Respir. Cell. Mol. Biol.*, 14:131-138, 1996). Numerous animal and human studies have already linked the progression of renal disease, especially its hallmark pathology of interstitial fibrosis and glomerular sclerosis, to increased signaling by TGF- $\beta$ 1. (August P, et al. *Curr. Hypertens. Rep.* 2:184-91, 2000).

10 Signaling by TGF- $\beta$ 1 involves specific binding of the ligand to the type II TGF- $\beta$ 1 receptor (abbreviated as TGF $\beta$ -RII), present on the plasma membrane of target cells such as fibroblasts in the case of glomerular and interstitial fibrosis. This receptor-ligand complex then heterodimerizes with the type I TGF- $\beta$ 1 receptor (abbreviated as TGF $\beta$ -RI). TGF $\beta$ -RI is constitutively active. Like the 15 concentrations of ligand (TGF- $\beta$ 1) and TGF $\beta$ -RI, the concentration of TGF $\beta$ -RII in the plasma membrane is likely to be rate-limiting for signaling by TGF- $\beta$ 1. All elements of the pathway appear to be subject to complex regulation. TGF- $\beta$ 1 signaling has been identified, and methods of developing therapies based on these regulatory reactions have been characterized (for example, see Souchelnytokyi, et 20 al., U.S. Pat No. 6,103,869, or Falb, U.S. Pat No. 6,099,823).

Activation of protein kinase C early during compensatory renal growth (CRF) would have the effect of stimulating TGF- $\beta$ 1 production, since the TGF- $\beta$ 1 promoter contains AP-1 sites (Kim et al., *J. Biol. Chem.*, 264:402-408, 1989). Angiotensin II has been shown to induce TGF- $\beta$ 1 expression in renal mesangial 25 cells, endothelial cells, and proximal tubular epithelial cells. Thus, greater induction of TGF- $\beta$ 1, or greater expression of its two main receptors (TGF $\beta$ -RI and TGF $\beta$ -RII), may occur in patients who progress to ESRD compared to patients who never develop CRF. Unlike the case with renal failure, TGF- $\beta$ 1 signaling has not been implicated in essential hypertension yet.

30 If the level of TGF $\beta$ -RII gene product (i.e. protein) is proportional to the level of mRNA, and the mRNA level is proportional to the transcriptional rate of the gene, then a SNP which disrupts a transcriptional activator site would be expected to

decrease both the rate of transcription of the gene and the eventual concentration of TGF $\beta$ -RII in the plasma membrane of cells which express this protein. The net effect of such a SNP is expected to be protection against renal failure.

Since the coding sequence of TGF- $\beta$ 1 is identical between mouse and 5 human, a period of evolutionary divergence of greater than 100 hundred million years, no human polymorphisms in the coding sequence are expected. Thus the TGF- $\beta$ 1 promoter and introns would be more likely candidates for genetic variants than the exons of the TGF- $\beta$ 1 structural gene. The promoter sequences and the structural genes for TGF $\beta$ -RI and TGF $\beta$ -RII are also likely candidates for genetic 10 variations.

Those of ordinary skill in the art will recognize that alterations in the regulatory region of a gene, i.e. promoter, can produce substantive changes in the timing and quantity of the production of said gene's product. GC box elements are a relatively common regulatory motif (2.12 matches/1000 bases of random genomic 15 DNA in vertebrates). Mutations in a GC box located at -90 of the **human**  $\beta$ -globin transcription startpoint result in suppression of transcription to as low as 10% of the normal level (Lewin, B. *Genes VII*; New York: Oxford University Press, 1999; pp. 634-635). If the level of TGF $\beta$ -RII gene product (i.e. protein) is proportional to the level of mRNA, and the mRNA level is proportional to the transcriptional rate of the 20 gene, then a SNP which disrupts a transcriptional activator site would be expected to decrease both the rate of transcription of the gene and the eventual concentration of TGF $\beta$ -RII in the plasma membrane of cells which express this protein. The net effect of such a SNP is expected to be protection against renal failure.

An ideal approach to prevention of ESRD would be the identification of any 25 genes that predispose an individual to ESRD early enough to be able to counteract this predisposition. Knowledge of ESRD-predisposing genes is essential for truly effective delay, or, ideally, prevention of ESRD.

## SUMMARY

The present inventor has discovered novel single nucleotide polymorphisms 30 (SNPs) associated with the development of hypertension and/or end-stage renal disease in patients with hypertension. As such, these polymorphisms provide a

method for diagnosing a genetic predisposition for the development of hypertension or end-stage renal disease in individuals. Information obtained from the detection of SNPs associated with the development of these diseases is of great value in the treatment and prevention of the diseases.

5 Accordingly, one aspect of the present invention provides a method for diagnosing a genetic predisposition for hypertension and/or end-stage renal disease in a subject, comprising obtaining a sample containing at least one polynucleotide from the subject, and analyzing at least the polynucleotide to detect a genetic polymorphism wherein said genetic polymorphism is associated with an altered 10 susceptibility to developing hypertension and/or end stage renal disease.

15 Another aspect of the present invention provides an isolated nucleic acid sequence comprising at least 10 contiguous nucleotides from SEQ ID NO: 1, or their complements, wherein the sequence contains at least one polymorphic site associated with a disease and in particular hypertension and/or end-stage renal disease.

20 Yet another aspect of the invention is a kit for the detection of a polymorphism comprising, at a minimum, at least one polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO: 1, or their complements, wherein the at least one polynucleotide contains at least one polymorphic site associated with hypertension and/or end-stage renal disease.

25 Yet another aspect of the invention provides a method for treating hypertension and/or end stage renal disease comprising, obtaining a sample of biological material containing at least one polynucleotide from the subject; analyzing the polynucleotide to detect the presence of at least one polymorphism associated with these diseases; and treating the subject in such a way as to 30 counteract the effect of any such polymorphism detected.

Still another aspect of the invention provides a method for the prophylactic treatment of a subject with a genetic predisposition to hypertension and/or end stage renal disease comprising, obtaining a sample of biological material containing at least one polynucleotide from the subject; analyzing the polynucleotide to detect the presence of at least one polymorphism associated with these diseases; and treating the subject.

Further scope of the applicability of the present invention will become apparent from the detailed description and drawings provided below. It should be understood, however, that the following detailed description and examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from the following detailed description.

### DEFINITIONS

nt = nucleotide

10 bp = base pair

kb = kilobase; 1000 base pairs

ESRD = end-stage renal disease

HTN = hypertension

NIDDM = noninsulin-dependent diabetes mellitus

15 CRF = chronic renal failure

T-GF = tubulo-glomerular feedback

CRG = compensatory renal growth

MODY = maturity-onset diabetes of the young

RFLP = restriction fragment length polymorphism

20 MASDA = multiplexed allele-specific diagnostic assay

MADGE = microtiter array diagonal gel electrophoresis

OLA = oligonucleotide ligation assay

DOL = dye-labeled oligonucleotide ligation assay

SNP = single nucleotide polymorphism

25 PCR = polymerase chain reaction

“polynucleotide” and “oligonucleotide” are used interchangeably and mean a linear polymer of at least 2 nucleotides joined together by phosphodiester bonds and may consist of either ribonucleotides or deoxyribonucleotides.

“sequence” means the linear order in which monomers occur in a polymer, 30 for example, the order of amino acids in a polypeptide or the order of nucleotides in a polynucleotide.

“polymorphism” refers to a set of genetic variants at a particular genetic locus among individuals in a population.

“promoter” means a regulatory sequence of DNA that is involved in the binding of RNA polymerase to initiate transcription of a gene. A “gene” is a 5 segment of DNA involved in producing a peptide, polypeptide, or protein, including the coding region, non-coding regions preceding (“leader”) and following (“trailer”) coding region, as well as intervening non-coding sequences (“introns”) between individual coding segments (“exons”). A promoter is herein considered as a part of the corresponding gene. Coding refers to the representation of amino acids, start and 10 stop signals in a three base “triplet” code. Promoters are often upstream (“5’ to”) the transcription initiation site of the gene.

“gene therapy” means the introduction of a functional gene or genes from some source by any suitable method into a living cell to correct for a genetic defect.

“wild type allele” means the most frequently encountered allele of a given 15 nucleotide sequence of an organism.

“genetic variant” or “variant” means a specific genetic variant which is present at a particular genetic locus in at least one individual in a population and that differs from the wild type.

As used herein the terms “patient” and “subject” are not limited to human 20 beings, but are intended to include all vertebrate animals in addition to human beings.

As used herein the terms “genetic predisposition”, “genetic susceptibility” and “susceptibility” all refer to the likelihood that an individual subject will develop a particular disease, condition or disorder. For example, a subject with an increased 25 susceptibility or predisposition will be more likely than average to develop a disease, while a subject with a decreased predisposition will be less likely than average to develop the disease. A genetic variant is associated with an altered susceptibility or predisposition if the allele frequency of the genetic variant in a population or subpopulation with a disease, condition or disorder varies from its allele frequency 30 in the population without the disease, condition or disorder (control population) or a control sequence (wild type) by at least 1%, preferably by at least 2%, more preferably by at least 4% and more preferably still by at least 8%.

As used herein "isolated nucleic acid" means a species of the invention that is the predominate species present (e.g., on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all 5 macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

As used herein, "allele frequency" means the frequency that a given allele appears in a population.

10

## DETAILED DESCRIPTION

All publications, patents, patent applications and other references cited in this application are herein incorporated by reference in their entirety as if each individual publication, patent, patent application or other reference were specifically and individually indicated to be incorporated by reference.

15

### Novel Polymorphisms

The present application provides six single nucleotide polymorphisms (SNPs) in genes associated with hypertension and/or end stage renal disease due to hypertension. The location of these SNPs associated with end stage renal disease as well as the wild type and variant nucleotides are summarized in Table 13. The 20 location of these SNPs associated with hypertension as well as the wild type and variant nucleotides are summarized in Table 14.

### Role of SNP-Typing

Because the complexity of transcription allows for factors of multiple functions to recognize the same regulatory elements, and the functional nature of 25 TGF  $\beta$  signaling is context-dependent, it is extraordinarily difficult to predict at this time the precise impact that natural genetic variation in these regions may have on human pathology. Therefore, the most immediate way to understand and benefit from the knowledge of this natural human variation is statistical analysis of diseased populations. Many statistical techniques exist for quantifying the association 30 between disease genes and disease phenotypes; the most robust for dissecting complex diseases, e.g. end-stage renal disease, is the case-control study design

(Risch, N. & Merikangas, K. *Science* 273, 1516-1517 (1996).)

Further, well-known genotyping techniques can be performed to type polymorphisms that are in close proximity to mutations in the target gene itself, including mutations associated with fibroproliferative, oncogenic or cardiovascular disorders. Such polymorphisms can be used to identify individuals of a population likely to carry mutations in the target gene e.g., TGF  $\beta$  type II receptor or a related gene. If a polymorphism exhibits linkage disequilibrium with mutations in the target gene e.g., TGF  $\beta$  type II receptor, the polymorphism can also be used to identify individuals in the general population who are likely to carry such mutations.

For example, Drazen et al. (U.S. Pat. No. 6,090,547) describe a technique using SSCP to detect substitution polymorphisms, and SSLP to detect insertion/deletion polymorphisms, in the coding and regulatory regions of the 5-lipoxygenase gene. Furthermore, they demonstrate that these polymorphisms can be usefully associated with asthmatic phenotypes, the knowledge of which is used to predict a response to conventional asthma therapy.

Also, Weber (U.S. Pat. No. 5,075,217) describes a DNA marker based on length (i.e. insertion/deletion) polymorphisms in blocks of (dC-dA) $n$ -(dG-dT) $n$  short tandem repeats. The average separation of (dC-dA) $n$ -(dG-dT) $n$  blocks is estimated to be 30,000-60,000 bp. Markers that are so closely spaced exhibit a high frequency co-inheritance, and are extremely useful in the identification of genetic mutations, such as, for example, mutations within TGF $\beta$ -RII or a related gene, and the diagnosis of diseases and disorders related to mutations in the target gene.

Also, Caskey et al. (U.S. Pat. No. 5,364,759) describe a DNA profiling assay for detecting short tri and tetra nucleotide repeat sequences. The process includes extracting the DNA of interest, such as the target gene, e.g., TGF $\beta$ -RII or a related gene, amplifying the extracted DNA, and labeling the repeat sequences to form a genotypic map of the individual's DNA.

For a further example of the use of genetic markers in disease diagnosis, see Shor, et al. U.S. Pat. No. 5,424,187.

### Preparation of Samples

The presence of genetic variants in the above genes or their control regions, or in any other genes that may affect susceptibility to ESRD is determined by

5 screening nucleic acid sequences from a population of individuals for such variants. The population is preferably comprised of some individuals with ESRD, so that any genetic variants that are found can be correlated with ESRD. The population is also preferably comprised of some individuals that have known risk for ESRD, such as individuals with hypertension, NIDDM, or CRF. The population should preferably

10 be large enough to have a reasonable chance of finding individuals with the sought-after genetic variant. As the size of the population increases, the ability to find significant correlations between a particular genetic variant and susceptibility to ESRD also increases. Preferably, the population should have 10 or more individuals.

15 The nucleic acid sequence can be DNA or RNA. For the assay of genomic DNA, virtually any biological sample containing genomic DNA (e.g. not pure red blood cells) can be used. For example, and without limitation, genomic DNA can be conveniently obtained from whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal cells, skin or hair. For assays using cDNA or mRNA, the target

20 nucleic acid must be obtained from cells or tissues that express the target sequence. One preferred source and quantity of DNA is 10 to 30 ml of anticoagulated whole blood, since enough DNA can be extracted from leukocytes in such a sample to perform many repetitions of the analysis contemplated herein.

Many of the methods described herein require the amplification of DNA

25 from target samples. This can be accomplished by any method known in the art but preferably is by the polymerase chain reaction (PCR). Optimization of conditions for conducting PCR must be determined for each reaction and can be accomplished without undue experimentation by one of ordinary skill in the art. In general, methods for conducting PCR can be found in U.S. Patent Nos 4,965,188, 4,800,159,

30 4,683,202, and 4,683,195; Ausbel et al., eds., *Short Protocols in Molecular Biology*, 3<sup>rd</sup> ed., Wiley, 1995; and Innis et al., eds., *PCR Protocols*, Academic Press, 1990.

Other amplification methods include the ligase chain reaction (LCR) (see, Wu and Wallace, *Genomics*, 4:560-569, 1989; Landegren et al., *Science*, 241:1077-1080, 1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA*, 86:1173-1177, 1989), self-sustained sequence replication (Guatelli et al., *Proc. Natl. Acad. Sci. USA*, 87:1874-1878, 1990), and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produces both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

## 10 **Detection of Polymorphisms**

### Detection of Unknown Polymorphisms

Two types of detection are contemplated within the present invention. The first type involves detection of unknown SNPs by comparing nucleotide target sequences from individuals in order to detect sites of polymorphism. If the most common sequence of the target nucleotide sequence is not known, it can be determined by analyzing individual humans, animals or plants with the greatest diversity possible. Additionally the frequency of sequences found in subpopulations characterized by such factors as geography or gender can be determined.

The presence of genetic variants and in particular SNPs is determined by screening the DNA and/or RNA of a population of individuals for such variants. If it is desired to detect variants associated with a particular disease or pathology, the population is preferably comprised of some individuals with the disease or pathology, so that any genetic variants that are found can be correlated with the disease of interest. It is also preferable that the population be composed of individuals with known risk factors for the disease. The populations should preferably be large enough to have a reasonable chance to find correlations between a particular genetic variant and susceptibility to the disease of interest. In one embodiment, the population should have at least 10 individuals, in another embodiment, the population should have 40 individuals or more. In one embodiment, the population is preferably comprised of individuals who have known risk factors for ESRD such as individuals with hypertension, NIDDM, or CRF. In

addition, the allele frequency of the genetic variant in a population or subpopulation with the disease or pathology should vary from its allele frequency in the population without the disease or pathology (control population) or the control sequence (wild type) by at least 1%, preferably by at least 2%, more preferably by at least 4% and 5 more preferably still by at least 8%.

Determination of unknown genetic variants, and in particular SNPs, within a particular nucleotide sequence among a population may be determined by any method known in the art, for example and without limitation, direct sequencing, restriction length fragment polymorphism (RFLP), single-strand conformational 10 analysis (SSCA), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis (HET), chemical cleavage analysis (CCM) and ribonuclease cleavage.

Methods for direct sequencing of nucleotide sequences are well known to those skilled in the art and can be found for example in Ausubel et al., eds., *Short Protocols in Molecular Biology*, 3<sup>rd</sup> ed., Wiley, 1995 and Sambrook et al., 15 *Molecular Cloning*, 2<sup>nd</sup> ed., Chap. 13, Cold Spring Harbor Laboratory Press, 1989. Sequencing can be carried out by any suitable method, for example, dideoxy sequencing (Sanger et al., *Proc. Natl. Acad. Sci. USA*, 74:5463-5467, 1977), chemical sequencing (Maxam and Gilbert, *Proc. Natl. Acad. Sci. USA*, 74:560-564, 1977) or variations thereof. Direct sequencing has the advantage of determining 20 variation in any base pair of a particular sequence.

In one embodiment, direct sequencing is accomplished by pyrosequencing. In pyrosequencing, a sequencing primer is hybridized with a DNA template and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates, adenosine 5' phosphosulfate (APS) and luciferin. The 25 first of four deoxynucleotide triphosphates (dNTP) is added to the reaction and incorporated into the DNA primer strand if it is complementary to the base in the template. Each dNTP incorporation is accompanied by release of pyrophosphate (PPi) in an quantity equimolar to the amount of incorporated nucleotide. ATP sulfurylase then quantitatively converts the PPi to ATP in the presence of adenosine 30 5' phosphosulfate. The ATP produced drives the luciferase mediated conversion of luciferin to oxyluciferin which generates visible light in amounts proportional to the amount of ATP. The amount of light produced is measured and is proportional to

the number of nucleotides incorporated. The reaction is then repeated for each of the remaining dNTPs. For dATP, alfa-thio triphosphate (dATPS) is used since it is efficiently utilized by DNA polymerase but not by luciferase. Methods for using pyrosequencing to detect SNPs are known in the art and can be found, for example, 5 in Alderborn et al., *Genome Res.* 10:1249-1258, 2000; Ahmadian et al., *Anal. Biochem.* 10:103-110, 2000; and Nordstrom et al., *Biotechnol. Appl. Biochem.* 31:107-112, 2000.

RFLP analysis (see, e.g. U.S. Patents No. 5,324,631 and 5,645,995) is useful for detecting the presence of genetic variants at a locus in a population when the 10 variants differ in the size of a probed restriction fragment within the locus, such that the difference between the variants can be visualized by electrophoresis. Such differences will occur when a variant creates or eliminates a restriction site within the probed fragment. RFLP analysis is also useful for detecting a large insertion or deletion within the probed fragment. Thus, RFLP analysis is useful for detecting, 15 e.g., an *Alu* sequence insertion or deletion in a probed DNA segment.

Single-strand conformational polymorphisms (SSCPs) can be detected in <220 bp PCR amplicons with high sensitivity (Orita et al, *Proc. Natl. Acad. Sci. USA*, 86:2766-2770, 1989; Warren et al., In: *Current Protocols in Human Genetics*, Dracopoli et al., eds, Wiley, 1994, 7.4.1-7.4.6.). Double strands are first heat- 20 denatured. The single strands are then subjected to polyacrylamide gel electrophoresis under non-denaturing conditions at constant temperature (i.e. low voltage and long run times) at two different temperatures, typically 4-10°C and 23°C (room temperature). At low temperatures (4-10°C), the secondary structure of short single strands (degree of intrachain hairpin formation) is sensitive to even single 25 nucleotide changes, and can be detected as a large change in electrophoretic mobility. The method is empirical, but highly reproducible, suggesting the existence of a very limited number of folding pathways for short DNA strands at the critical temperature. Polymorphisms appear as new banding patterns when the gel is stained.

Denaturing gradient gel electrophoresis (DGGE) can detect single base 30 mutations based on differences in migration between homo- and heteroduplexes (Myers et al., *Nature*, 313:495-498, 1985). The DNA sample to be tested is hybridized to a labeled wild type probe. The duplexes formed are then subjected to

electrophoresis through a polyacrylamide gel that contains a gradient of DNA denaturant parallel to the direction of electrophoresis. Heteroduplexes formed due to single base variations are detected on the basis of differences in migration between the heteroduplexes and the homoduplexes formed.

5 In heteroduplex analysis (HET) (Keen et al., *Trends Genet.* 7:5, 1991), genomic DNA is amplified by the polymerase chain reaction followed by an additional denaturing step which increases the chance of heteroduplex formation in heterozygous individuals. The PCR products are then separated on Hydrolink gels where the presence of the heteroduplex is observed as an additional band.

10 Chemical cleavage analysis (CCM) is based on the chemical reactivity of thymine (T) when mismatched with cytosine, guanine or thymine and the chemical reactivity of cytosine (C) when mismatched with thymine, adenine or cytosine (Cotton et al., *Proc. Natl. Acad. Sci. USA*, 85:4397-4401, 1988). Duplex DNA formed by hybridization of a wild type probe with the DNA to be examined, is  
15 treated with osmium tetroxide for T and C mismatches and hydroxylamine for C mismatches. T and C mismatched bases that have reacted with the hydroxylamine or osmium tetroxide are then cleaved with piperidine. The cleavage products are then analyzed by gel electrophoresis.

20 Ribonuclease cleavage involves enzymatic cleavage of RNA at a single base mismatch in an RNA:DNA hybrid (Myers et al., *Science* 230:1242-1246, 1985). A <sup>32</sup>P labeled RNA probe complementary to the wild type DNA is annealed to the test DNA and then treated with ribonuclease A. If a mismatch occurs, ribonuclease A will cleave the RNA probe and the location of the mismatch can then be determined by size analysis of the cleavage products following gel electrophoresis.

25 Detection of Known Polymorphisms

20 The second type of polymorphism detection involves determining which form of a known polymorphism is present in individuals for diagnostic or epidemiological purposes. In addition to the already discussed methods for detection of polymorphisms, several methods have been developed to detect known  
30 SNPs. Many of these assays have been reviewed by Landegren et al., *Genome Res.*, 8:769-776, 1998, and will only be briefly reviewed here.

One type of assay has been termed an array hybridization assay, an example of which is the multiplexed allele-specific diagnostic assay (MASDA) (U.S. Patent No. 5,834,181; Shuber et al., *Hum. Molec. Genet.*, 6:337-347, 1997). In MASDA, samples from multiplex PCR are immobilized on a solid support. A single 5 hybridization is conducted with a pool of labeled allele specific oligonucleotides (ASO). Any ASOs that hybridize to the samples are removed from the pool of ASOs. The support is then washed to remove unhybridized ASOs remaining in the pool. Labeled ASOs remaining on the support are detected and eluted from the support. The eluted ASOs are then sequenced to determine the mutation present.

10 Two assays depend on hybridization-based allele-discrimination during PCR. The TaqMan assay (U.S. Patent No. 5,962,233; Livak et al., *Nature Genet.*, 9:341-342, 1995) uses allele specific (ASO) probes with a donor dye on one end and an acceptor dye on the other end, such that the dye pair interact via fluorescence resonance energy transfer (FRET). A target sequence is amplified by PCR modified 15 to include the addition of the labeled ASO probe. The PCR conditions are adjusted so that a single nucleotide difference will effect binding of the probe. Due to the 5' nuclease activity of the *Taq* polymerase enzyme, a perfectly complementary probe is cleaved during the PCR while a probe with a single mismatched base is not cleaved. Cleavage of the probe dissociates the donor dye from the quenching acceptor dye, 20 greatly increasing the donor fluorescence.

An alternative to the TaqMan assay is the molecular beacons assay (U.S. Patent No. 5,925,517; Tyagi et al., *Nature Biotech.*, 16:49-53, 1998). In the molecular beacons assay, the ASO probes contain complementary sequences flanking the target specific species so that a hairpin structure is formed. The loop of 25 the hairpin is complimentary to the target sequence while each arm of the hairpin contains either donor or acceptor dyes. When not hybridized to a donor sequence, the hairpin structure brings the donor and acceptor dye close together thereby extinguishing the donor fluorescence. When hybridized to the specific target sequence, however, the donor and acceptor dyes are separated with an increase in 30 fluorescence of up to 900 fold. Molecular beacons can be used in conjunction with amplification of the target sequence by PCR and provide a method for real time detection of the presence of target sequences or can be used after amplification.

High throughput screening for SNPs that affect restriction sites can be achieved by Microtiter Array Diagonal Gel Electrophoresis (MADGE) (Day and Humphries, *Anal. Biochem.*, 222:389-395, 1994). In this assay restriction fragment digested PCR products are loaded onto stackable horizontal gels with the wells arrayed in a microtiter format. During electrophoresis, the electric field is applied at an angle relative to the columns and rows of the wells allowing products from a large number of reactions to be resolved.

Additional assays for SNPs depend on mismatch distinction by polymerases and ligases. The polymerization step in PCR places high stringency requirements on correct base pairing of the 3' end of the hybridizing primers. This has allowed the use of PCR for the rapid detection of single base changes in DNA by using specifically designed oligonucleotides in a method variously called PCR amplification of specific alleles (PASA) (Sommer et al., *Mayo Clin. Proc.*, 64:1361-1372 1989; Sarker et al., *Anal. Biochem.* 1990), allele-specific amplification (ASA), allele-specific PCR, and amplification refractory mutation system (ARMS) (Newton et al., *Nuc. Acids Res.*, 1989; Nichols et al., *Genomics*, 1989; Wu et al., *Proc. Natl. Acad. Sci. USA*, 1989). In these methods, an oligonucleotide primer is designed that perfectly matches one allele but mismatches the other allele at or near the 3' end. This results in the preferential amplification of one allele over the other. By using three primers that produce two differently sized products, it can be determined whether an individual is homozygous or heterozygous for the mutation (Dutton and Sommer, *BioTechniques*, 11:700-702, 1991). In another method, termed bi-PASA, four primers are used; two outer primers that bind at different distances from the site of the SNP and two allele specific inner primers (Liu et al., *Genome Res.*, 7:389-398, 1997). Each of the inner primers has a non-complementary 5' end and form a mismatch near the 3' end if the proper allele is not present. Using this system, zygosity is determined based on the size and number of PCR products produced.

The joining by DNA ligases of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. This sensitivity has been utilized in the oligonucleotide ligation assay (Landegren et al., *Science*, 241:1077-1080, 1988) and the ligase chain reaction (LCR; Barany, *Proc. Natl. Acad. Sci. USA*, 88:189-193, 1991). In OLA, the

sequence surrounding the SNP is first amplified by PCR, whereas in LCR, genomic DNA can be used as a template.

In one method for mass screening for SNPs based on the OLA, amplified DNA templates are analyzed for their ability to serve as templates for ligation reactions between labeled oligonucleotide probes (Samotiaki et al., *Genomics*, 20:238-242, 1994). In this assay, two allele-specific probes labeled with either of two lanthanide labels (europium or terbium) compete for ligation to a third biotin labeled phosphorylated oligonucleotide and the signals from the allele specific oligonucleotides are compared by time-resolved fluorescence. After ligation, the oligonucleotides are collected on an avidin-coated 96-pin capture manifold. The collected oligonucleotides are then transferred to microtiter wells in which the europium and terbium ions are released. The fluorescence from the europium ions is determined for each well, followed by measurement of the terbium fluorescence.

In alternative gel-based OLA assays, numerous SNPs can be detected simultaneously using multiplex PCR and multiplex ligation (U.S. Patent No. 5,830,711; Day et al., *Genomics*, 29:152-162, 1995; Grossman et al., *Nuc. Acids Res.*, 22:4527-4534, 1994). In these assays, allele specific oligonucleotides with different markers, for example, fluorescent dyes, are used. The ligation products are then analyzed together by electrophoresis on an automatic DNA sequencer distinguishing markers by size and alleles by fluorescence. In the assay by Grossman et al., 1994, mobility is further modified by the presence of a non-nucleotide mobility modifier on one of the oligonucleotides.

A further modification of the ligation assay has been termed the dye-labeled oligonucleotide ligation (DOL) assay (U.S. Patent No. 5,945,283; Chen et al., *Genome Res.*, 8:549-556, 1998). DOL combines PCR and the oligonucleotide ligation reaction in a two-stage thermal cycling sequence with fluorescence resonance energy transfer (FRET) detection. In the assay, labeled ligation oligonucleotides are designed to have annealing temperatures lower than those of the amplification primers. After amplification, the temperature is lowered to a temperature where the ligation oligonucleotides can anneal and be ligated together. This assay requires the use of a thermostable ligase and a thermostable DNA polymerase without 5' nuclease activity. Because FRET occurs only when the

donor and acceptor dyes are in close proximity, ligation is inferred by the change in fluorescence.

In another method for the detection of SNPs termed minisequencing, the target-dependent addition by a polymerase of a specific nucleotide immediately downstream (3') to a single primer is used to determine which allele is present (U.S. Patent No. 5,846,710). Using this method, several SNPs can be analyzed in parallel by separating locus specific primers on the basis of size via electrophoresis and determining allele specific incorporation using labeled nucleotides.

Determination of individual SNPs using solid phase minisequencing has been described by Syvanen et al., *Am. J. Hum. Genet.*, 52:46-59, 1993. In this method the sequence including the polymorphic site is amplified by PCR using one amplification primer which is biotinylated on its 5' end. The biotinylated PCR products are captured in streptavidin-coated microtitration wells, the wells washed, and the captured PCR products denatured. A sequencing primer is then added whose 3' end binds immediately prior to the polymorphic site, and the primer is elongated by a DNA polymerase with one single labeled dNTP complementary to the nucleotide at the polymorphic site. After the elongation reaction, the sequencing primer is released and the presence of the labeled nucleotide detected. Alternatively, dye labeled dideoxynucleoside triphosphates (ddNTPs) can be used in the elongation reaction (U.S. Patent No. 5,888,819; Shumaker et al., *Human Mut.*, 7:346-354, 1996). In this method, incorporation of the ddNTP is determined using an automatic gel sequencer.

Minisequencing has also been adapted for use with microarrays (Shumaker et al., *Human Mut.*, 7:346-354, 1996). In this case, elongation (extension) primers are attached to a solid support such as a glass slide. Methods for construction of oligonucleotide arrays are well known to those of ordinary skill in the art and can be found, for example, in *Nature Genetics*, Suppl., Vol. 21, January, 1999. PCR products are spotted on the array and allowed to anneal. The extension (elongation) reaction is carried out using a polymerase, a labeled dNTP and noncompeting ddNTPs. Incorporation of the labeled dNTP is then detected by the appropriate means. In a variation of this method suitable for use with multiplex PCR, extension

is accomplished with the use of the appropriate labeled ddNTP and unlabeled ddNTPs (Pastinen et al., *Genome Res.*, 7:606-614, 1997).

Solid phase minisequencing has also been used to detect multiple polymorphic nucleotides from different templates in an undivided sample (Pastinen et al., *Clin. Chem.*, 42:1391-1397, 1996). In this method, biotinylated PCR products are captured on the avidin-coated manifold support and rendered single stranded by alkaline treatment. The manifold is then placed serially in four reaction mixtures containing extension primers of varying lengths, a DNA polymerase and a labeled ddNTP, and the extension reaction allowed to proceed. The manifolds are inserted into the slots of a gel containing formamide which releases the extended primers from the template. The extended primers are then identified by size and fluorescence on a sequencing instrument.

Fluorescence resonance energy transfer (FRET) has been used in combination with minisequencing to detect SNPs (U.S. Patent No. 5,945,283; Chen et al., *Proc. Natl. Acad. Sci. USA*, 94:10756-10761, 1997). In this method, the extension primers are labeled with a fluorescent dye, for example fluorescein. The ddNTPs used in primer extension are labeled with an appropriate FRET dye. Incorporation of the ddNTPs is determined by changes in fluorescence intensities.

The above discussion of methods for the detection of SNPs is exemplary only and is not intended to be exhaustive. Those of ordinary skill in the art will be able to envision other methods for detection of SNPs that are within the scope and spirit of the present invention.

In one embodiment the present invention provides a method for diagnosing a genetic predisposition for a disease and in particular, end-stage renal disease and hypertension. In this method, a biological sample is obtained from a subject. The subject can be a human being or any vertebrate animal. The biological sample must contain polynucleotides and preferably genomic DNA. Samples that do not contain genomic DNA, for example, pure samples of mammalian red blood cells, are not suitable for use in the method. The form of the polynucleotide is not critically important such that the use of DNA, cDNA, RNA or mRNA is contemplated within the scope of the method. The polynucleotide is then analyzed to detect the presence of a genetic variant where such variant is associated with an altered susceptibility to

a disease, condition or disorder, and in particular end-stage renal disease. In one embodiment, the genetic variant is located at one of the polymorphic sites contained in Table 13 or 14. In another embodiment, the genetic variant is one of the variants contained in Table 13 or 14 or the complement of any of the variants contained in Table 13 or 14. Any method capable of detecting a genetic variant, including any of the methods previously discussed, can be used. Suitable methods include, but are not limited to, those methods based on sequencing, mini sequencing, hybridization, restriction fragment analysis, oligonucleotide ligation, or allele specific PCR.

The present invention is also directed to an isolated nucleic acid sequence of at least 10 contiguous nucleotides from SEQ ID NO: 1, or the complement of SEQ ID NO: 1. In one preferred embodiment, the sequence contains at least one polymorphic site associated with a disease, and in particular end-stage renal disease. In one embodiment, the polymorphic site is selected from the groups contained in Table 13 or 14. In another embodiment, the polymorphic site contains a genetic variant, and in particular, the genetic variants contained in Table 13 or 14 or the complements of the variants in Table 13 or 14. In yet another embodiment, the polymorphic site, which may or may not also include a genetic variant, is located at the 3' end of the polynucleotide. In still another embodiment, the polynucleotide further contains a detectable marker. Suitable markers include, but are not limited to, radioactive labels, such as radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.

The present invention also includes kits for the detection of polymorphisms associated with diseases, conditions or disorders, and in particular end-stage renal disease and hypertension. The kits contain, at a minimum, at least one polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO 1, or the complement of SEQ ID NO: 1. In one embodiment, the polynucleotide contains at least one polymorphic site, preferably a polymorphic site selected from the groups contained in Table 13 or 14. Alternatively the 3= end of the polynucleotide is immediately 5' to a polymorphic site, preferably a polymorphic site contained in Table 13 or 14. In one embodiment, the polymorphic site contains a genetic variant, preferably a genetic variant selected from the groups contained in Table 13 or 14. In still another embodiment, the genetic variant is located at the 3= end of the

polynucleotide. In yet another embodiment, the polynucleotide of the kit contains a detectable label. Suitable labels include, but are not limited to, radioactive labels, such as radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.

5 In addition, the kit may also contain additional materials for detection of the polymorphisms. For example, and without limitation, the kits may contain buffer solutions, enzymes, nucleotide triphosphates, and other reagents and materials necessary for the detection of genetic polymorphisms. Additionally, the kits may contain instructions for conducting analyses of samples for the presence of 10 polymorphisms and for interpreting the results obtained.

In yet another embodiment the present invention provides a method for designing a treatment regime for a patient having a disease, condition or disorder and in particular end stage renal disease and hypertension caused either directly or indirectly by the presence of one or more single nucleotide polymorphisms. In this 15 method, genetic material from a patient, for example, DNA, cDNA, RNA or mRNA is screened for the presence of one or more SNPs associated with the disease of interest. Depending on the type and location of the SNP, a treatment regime is designed to counteract the effect of the SNP.

Alternatively, information gained from analyzing genetic material for the 20 presence of polymorphisms can be used to design treatment regimes involving gene therapy. For example, detection of a polymorphism that either affects the expression of a gene or results in the production of a mutant protein can be used to design an artificial gene to aid in the production of normal, wild type protein or help restore normal gene expression. Methods for the construction of polynucleotide sequences 25 encoding proteins and their associated regulatory elements are well known to those of ordinary skill in the art. Once designed, the gene can be placed in the individual by any suitable means known in the art (*Gene Therapy Technologies, Applications and Regulations*, Meager, ed., Wiley, 1999; *Gene Therapy: Principles and Applications*, Blankenstein, ed., Birkhauser Verlag, 1999; Jain, *Textbook of Gene Therapy*, 30 Hogrefe and Huber, 1998).

The present invention is also useful in designing prophylactic treatment regimes for patients determined to have an increased susceptibility to a disease,

condition or disorder, and in particular end stage renal disease and hypertension due to the presence of one or more single nucleotide polymorphisms. In this embodiment, genetic material, such as DNA, cDNA, RNA or mRNA, is obtained from a patient and screened for the presence of one or more SNPs associated either 5 directly or indirectly to a disease, condition, disorder or other pathological condition. Based on this information, a treatment regime can be designed to decrease the risk of the patient developing the disease. Such treatment can include, but is not limited to, surgery, the administration of pharmaceutical compounds or nutritional supplements, and behavioral changes such as improved diet, increased exercise, 10 reduced alcohol intake, smoking cessation, etc.

## EXAMPLES

Position of the single nucleotide polymorphism (SNP) is given according to the numbering scheme in GenBank Accession Number U37070. Thus, all nucleotides will be positively numbered, rather than bear negative numbers 15 reflecting their position upstream from the transcription initiation site, a scheme often used for promoters. The two numbering systems can be easily interconverted, if necessary. GenBank sequences can be found at <http://www.ncbi.nlm.nih.gov/>

In the following examples, SNPs are written as "reference sequence nucleotide" → "variant nucleotide" Changes in nucleotide sequences are indicated 20 in bold print. The standard nucleotide abbreviations are used in which A=adenine, C=cytosine, G=guanine, T=thymine, M=A or C, R=A or G, W=A or T, S=C or G, Y=C or T, K=G or T, V=A or C or G, H=A or C or T; D=A or G or T; B=C or G or T; N=A or C or G or T.

### Example 1

25 Detection of Novel Polymorphisms by Direct Sequencing of Leukocyte Genomic DNA

Leukocytes were obtained from human whole blood collected with EDTA. Blood was obtained from a group of 20 Caucasian males with ESRD due to hypertension, 23 Caucasian males with hypertension, and a control group of 29 30 Caucasian males.

Genomic DNA was purified from the collected leukocytes using standard protocols well known to those of ordinary skill in the art of molecular biology (Ausubel et al., *Short Protocol in Molecular Biology*, 3<sup>rd</sup> ed., John Wiley and Sons, 1995; Sambrook et al., *Molecular Cloning*, Cold Spring Harbor Laboratory Press, 1989; and Davis et al., *Basic Methods in Molecular Biology*, Elsevier Science Publishing, 1986). One hundred nanograms of purified genomic DNA was used in each PCR reaction.

Standard PCR reaction conditions were used. Methods for conducting PCR are well known in the art and can be found, for example, in U.S. Patent Nos 10 4,965,188, 4,800,159, 4,683,202, and 4,683,195; Ausbel et al., eds., *Short Protocols in Molecular Biology*, 3<sup>rd</sup> ed., Wiley, 1995; and Innis et al., eds., *PCR Protocols*, Academic Press, 1990. Specific primers used are given in the following examples.

PCR reactions were carried out in a total volume of 50  $\mu$ l containing 10-15 ng leukocyte genomic DNA, 10 pmol of each primer, 200 nM deoxynucleotide 15 triphosphates (dNTPs), 1.25 U Taq polymerase (Qiagen), 1X Qiagen PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl<sub>2</sub>, and 1X "Q" solution (Qiagen). After an initial 3 minutes denaturation at 94°C, 35 cycles were performed consisting of 1 minute denaturation at 94°C, 1 minute hybridization at 55°C, 2 minute extension at 72°C, followed by a final extension step of 5 minutes at 72°C, and 1 20 minute cooling at 35°C.

Post-PCR clean-up was performed as follows. PCR reactions were cleaned to remove unwanted primer and other impurities such as salts, enzymes, and unincorporated nucleotides that could inhibit sequencing. One of the following clean-up kits was used: Qiaquick-96 PCR Purification Kit (Qiagen) or Multiscreen- 25 PCR Plates (Millipore, discussed below).

When using the Qiaquick protocol, PCR samples were added to the 96-well Qiaquick silica-gel membrane plate and a chaotropic salt, supplied as "PB Buffer," was then added to each well. The PB Buffer causes DNA to bind to the membrane. The plate was put onto the Qiagen vacuum manifold and vacuum was applied to the 30 plate in order to pull sample and PB Buffer through the membrane. The filtrate was discarded. Next, the samples were washed twice using "PE Buffer." Vacuum pressure was applied between each step to remove the buffer. Filtrate was similarly

discarded after each wash. After the last PE Buffer wash, maximum vacuum pressure was applied to the membrane plate to generate maximum airflow through the membrane in order to evaporate residual ethanol left from the PE Buffer. The clean PCR product was then eluted from the filter using "EB Buffer." The filtrate 5 contained the cleaned PCR product and was collected. All buffers were supplied as part of the Qiaquick-96 PCR Purification Kit. The vacuum manifold was also purchased from Qiagen for exclusive use with the Qiaquick-96 Purification Kit.

When using the Millipore Multiscreen-PCR Plates, PCR samples were loaded into the wells of the Multiscreen-PCR Plate and the plate was then placed on 10 a Millipore vacuum manifold. Vacuum pressure was applied for 10 minutes, and the filtrate was discarded. The plate was then removed from the vacuum manifold and 100  $\mu$ l of Milli-Q water was added to each well to rehydrate the DNA samples. After shaking on a plate shaker for 5 minutes, the plate was replaced on the manifold and vacuum pressure was applied for 5 minutes. The filtrate was again discarded. 15 The plate was removed and 60  $\mu$ l Milli-Q water was added to each well to again rehydrate the DNA samples. After shaking on a plate shaker for 10 minutes, the 60  $\mu$ l of cleaned PCR product was transferred from the Multiscreen-PCR plate to another 96-well plate by pipetting. The Millipore vacuum manifold was purchased from Millipore for exclusive use with the Multiscreen-PCR plates.

20 Cycle sequencing was performed on the clean PCR product using an ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer). For a total volume of 20  $\mu$ l, the following reagents were added to each well of a 96-well plate: 2.0  $\mu$ l Terminator Ready Reaction mix, 3.0  $\mu$ l 5X Sequencing Buffer (ABI), 5-10  $\mu$ l template (30-90 ng double stranded DNA), 3.2 pM primer (primer 25 used was the forward primer from the PCR reaction), and Milli-Q water to 20  $\mu$ l total volume. The reaction plate was placed into a Hybaid thermal cycler block and programmed as follows: X 1 cycle: 1 degree/sec thermal ramp to 94°C, 94°C for 1 min; X 35 cycles: 1 degree/sec thermal ramp to 94°C, then 94°C for 10 sec, followed by 1 degree/sec thermal ramp to 50°C, then 50°C for 10 sec, followed by 1 degree/sec thermal ramp to 60°C, then 60°C for 4 minutes.

30 The cycle sequencing reaction product was cleaned up to remove the unincorporated dye-labeled terminators that can obscure data at the beginning of the

sequence. A precipitation protocol was used. To each sequencing reaction in the 96-well plate 20  $\mu$ l of Milli-Q water and 60  $\mu$ l of 100% isopropanol was added. The plate was left at room temperature for at least 20 minutes to precipitate the extension products. The plate was spun in a plate centrifuge (Jouan) at 3,000 x g for 30

5 minutes.

Without disturbing the pellet, the supernatant was discarded by inverting the plate onto several paper tissues (Kimiwipes) folded to the size of the plate. The inverted plate, with Kimiowipes in place, was placed into the centrifuge (Jouan) and spun at 700 x g for 1 minute. The Kimiowipes were discarded and the samples were

10 loaded onto a sequencing gel.

Approximately 1  $\mu$ l of sequencing product was loaded into each well of a 96-lane 5% Long Ranger (FMC single pack) gel. The running buffer consisted of 1X TBE (Tris Borate EDTA). The glass plates consisted of ABI 48-cm plates for use with a 96-lane 0.4 mm Mylar shark-tooth comb. A semi-automated ABI Prism 377-15 96 DNA sequencer was used (ABI 377 with 96-lane, Big Dye upgrades).

Sequencing run settings were as follows: run module 48E-1200, 8 hr collection time, 2400 V electrophoresis voltage, 50 mA electrophoresis current, 200 W electrophoresis power, CCD offset of 0, gel temperature of 51°C, 40 mW laser power, and CCD gain of 2.

20

### Example 2

#### A to C Substitution at Position 796 of Human TGF $\beta$ -RII Promoter

Table 1

| ALLELLE FREQUENCIES                        |     |      |
|--|-----|------|
|  | A   | C    |
| <u>CONTROL</u> (n=58 chromosomes):         | 43  | 15   |
| Caucasian men                              | 74% | 26%  |
| <u>DISEASE</u>                             |     |      |
| <u>HYPERTENSION</u> (n=46 chromosomes):    | 41  | 5    |
| Caucasian men                              | 89% | 11%  |
| <u>ESRD due to HTN</u> (n=40 chromosomes): | 39  | 1    |
| Caucasian men                              | 98% | 2.5% |

Table 2

| GENOTYPE FREQUENCIES                       |            |            |            |
|--|------------|------------|------------|
|  | <u>A/A</u> | <u>A/C</u> | <u>C/C</u> |
| <u>CONTROL</u> (n=29 individuals):         | 14         | 15         | 0          |
| Caucasian men                              | 48%        | 52%        | 0%         |
| <u>DISEASE</u>                             |            |            |            |
| <u>HYPERTENSION</u> (n=23 individuals):    | 18         | 5          | 0          |
| Caucasian men                              | 78%        | 22%        | 0%         |
| <u>ESRD due to HTN</u> (n=20 individuals): | 19         | 1          | 0          |
| Caucasian men                              | 95%        | 5%         | 0%         |

PCR and sequencing were conducted as in Example 1. The sense primer was 5'–GGAGTTGGGTTGGGGAG-3' (SEQ ID NO: 2) and the anti-sense primer was 5'–TCTTGCTAGGGCAACCAGATTG-3' (SEQ ID NO: 3). The PCR product spanned bases 697 to 988 of the TGF- $\beta$ -RII promoter (SEQ ID NO: 1).

As demonstrated above, the frequency of the C allele in Caucasian male hypertensive patients is less than half of that of the control sample of white men, 11% vs. 26%. The frequency of the C allele is over ten times lower in a sample of Caucasian male patients with ESRD due to hypertension compared to controls, 2.5% vs. 26%. The genotype frequencies are also dramatic: the frequency of the A/C genotype decreases over two-fold from control (52%) to hypertensive white male patients (22%), and over ten-fold to white men with ESRD due to hypertension (5%).

These data indicate that the reference sequence "A" allele contributes significantly towards hypertension and even more significantly towards ESRD as a complication of hypertension. Put differently, the C allele, i.e. the SNP at this position, appears to be strongly protective against hypertension and even more strongly protective against ESRD as a complication of hypertension.

These data roughly satisfy Hardy-Weinberg equilibrium for the control sample. A frequency of 0.74 for the A allele ("p") and 0.26 for the C allele ("q") among control individuals (see "Allele Frequencies," above) predicts genotype

frequencies of 55% A/A, 38% A/C, and 7% C/C at Hardy-Weinberg equilibrium ( $p^2 + 2pq + q^2 = 1$ ). The observed genotype frequencies were 48% A/A, 52% A/C, and 0% G/G, in rough agreement with those predicted for Hardy-Weinberg equilibrium.

The fact that the two disease categories diverge greatly from Hardy-  
5 Weinberg equilibrium is consistent with the hypothesis that this SNP is truly  
disease-associated.

The A796->C SNP is predicted to have a negative effect on transcription of  
the TGF $\beta$ -RII gene by disrupting a potential TCF11 (TCF11/KCR-F1/Nrf1  
homodimer) binding site beginning at nucleotide 788 on the (+) strand. The binding  
10 site consists of the sequence 5'-GTCATNN*W*NNNNNN-3' (SEQ ID NO: 4). This  
SNP replaces the underlined *W* (A or T) with a C. TCF11 homodimer sites occur  
relatively frequently, 4.63 matches per 1000 base pairs of random genomic sequence  
in vertebrates.

The TCF11 homodimer is a transcriptional activator, so disruption of its  
15 binding site in the TGF $\beta$ -RII promoter is expected to result in a lower rate of TGF $\beta$ -  
RII transcription, and a lower rate of TGF- $\beta$ 1 signaling, as discussed above. The  
A796->C SNP is therefore expected to be protective for the development of renal  
failure, since the currently accepted model of progression of chronic renal failure  
involves increased TGF- $\beta$ 1 signaling. These data are in full agreement. Among  
20 patients with end-stage renal disease, the A/A genotype (95%) is present almost  
twice as often as in the control population (48%; see "Genotype Frequencies,"  
above).

It is interesting that patients with hypertension but no renal failure have an  
intermediate frequency of the protective A796->C SNP, suggesting that  
25 hypertension itself may be due to increased TGF- $\beta$ 1 signaling. Such a mechanism  
would be novel.

From the standpoint of both molecular epidemiology and molecular genetics  
as discussed above, the A796->C SNP appears to be very important for  
hypertension, and even more important for ESRD due to hypertension.

## Example 3

A to C Substitution at Position 820 of Human TGF $\beta$ -RII Promoter

Table 3

| ALLELLE FREQUENCIES                 |          |          |
|-------------------------------------|----------|----------|
|                                     | <u>A</u> | <u>C</u> |
| <u>CONTROL</u> (n=58 chromosomes):  | 44       | 14       |
| Caucasian men                       | 76%      | 24%      |
| <u>DISEASE</u>                      |          |          |
| HYPERTENSION (n=46 chromosomes):    | 35       | 11       |
| Caucasian men                       | 76%      | 24%      |
| ESRD due to HTN (n=40 chromosomes): | 39       | 1        |
| Caucasian men                       | 98%      | 2.5%     |

5 Table 4

| GENOYPE FREQUENCIES                 |            |            |            |
|-------------------------------------|------------|------------|------------|
|                                     | <u>A/A</u> | <u>A/C</u> | <u>C/C</u> |
| <u>CONTROL</u> (n=29 individuals):  | 15         | 14         | 0          |
| Caucasian men                       | 52%        | 48%        | 0%         |
| <u>DISEASE</u>                      |            |            |            |
| HYPERTENSION (n=23 individuals):    | 12         | 11         | 0          |
| Caucasian men                       | 52%        | 48%        | 0%         |
| ESRD due to HTN (n=20 individuals): | 19         | 1          | 0          |
| Caucasian men                       | 95%        | 5%         | 0%         |

PCR and sequencing were conducted as in Example 1. The primers used were the same as in Example 2. The frequency of the C allele is almost ten times lower (2.5% vs. 24%) in white men with ESRD due to hypertension compared to a 10 control sample of white men. The frequency of the C allele among white men with hypertension, but without renal failure, is the same as the control group. The

genotype frequencies are equally dramatic: the frequency of the A/C genotype decreases ten-fold from control (48%) and hypertension (48%) groups to only 5% for white men with ESRD due to hypertension.

These data roughly satisfy the Hardy-Weinberg equilibrium for the control 5 sample. A frequency of 0.76 for the A allele ("p") and 0.24 for the C allele ("q") among control individuals predicts genotype frequencies of 58% A/A, 36% A/C, and 6% C/C at Hardy-Weinberg equilibrium ( $p^2 + 2pq + q^2 = 1$ ). The observed genotype frequencies were 52% A/A, 48% A/C, and 0% G/G, in rough agreement with those predicted for Hardy-Weinberg equilibrium.

10 The ESRD sample, but not the essential hypertension sample, diverges greatly from Hardy-Weinberg equilibrium, consistent with the hypothesis that this SNP is associated with ESRD but not hypertension.

The A820->C SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene by disrupting a potential TCF11 (TCF11/KCR-F1/Nrf1 homodimer) 15 binding site beginning at nucleotide 788 on the (+) strand of the TGF $\beta$ -RII promoter. The binding site consists of the sequence 5'-GTCATNNWNNNNNN-3'(SEQ ID NO: 5). This SNP replaces the underlined *W* (A or T) with a C. TCF11 homodimer sites occur relatively frequently, 4.63 matches per 1000 base pairs of random genomic sequence in vertebrates.

20 The TCF11 homodimer is a transcriptional activator, so disruption of its binding site in the TGF $\beta$ -RII promoter is expected to result in a lower rate of TGF $\beta$ -RII transcription, and a lower rate of TGF- $\beta$ 1 signaling. The A820->C SNP is therefore expected to be protective for the development of renal failure, since the currently accepted model of progression of chronic renal failure involves increased 25 TGF- $\beta$ 1 signaling. These data are in full agreement with such a model. Among patients with end-stage renal disease, the A/A genotype (95%) is present almost twice as often as in the control population (52%).

From the standpoint of both molecular epidemiology and molecular genetics, 30 the A820->C SNP appears to be associated with ESRD due to hypertension. These data indicate that the reference sequence "A" allele contributes significantly towards ESRD as a complication of hypertension. Put differently, the C allele, i.e. the single nucleotide polymorphism at this position, appears to be strongly protective against

ESRD as a complication of hypertension.

**Example 4**

C to G Substitution at Position 845 of Human TGF $\beta$ -RII Promoter

Table 5

| ALLELLE FREQUENCIES                        |          |          |
|--|----------|----------|
|  | <u>C</u> | <u>G</u> |
| <u>CONTROL</u> (n=58 chromosomes):         | 44       | 14       |
| Caucasian men                              | 76%      | 24%      |
| <u>DISEASE</u>                             |          |          |
| <u>HYPERTENSION</u> (n=46 chromosomes):    | 32       | 14       |
| Caucasian men                              | 70%      | 30%      |
| <u>ESRD due to HTN</u> (n=40 chromosomes): | 35       | 5        |
| Caucasian men                              | 88%      | 13%      |

5

Table 6

| GENOTYPE FREQUENCIES                       |            |            |            |
|--|------------|------------|------------|
|  | <u>C/C</u> | <u>C/G</u> | <u>G/G</u> |
| <u>CONTROL</u> (n=29 individuals):         | 16         | 12         | 1          |
| Caucasian men                              | 55%        | 41%        | 3%         |
| <u>DISEASE</u>                             |            |            |            |
| <u>HYPERTENSION</u> (n=23 individuals):    | 11         | 10         | 2          |
| Caucasian men                              | 48%        | 43%        | 9%         |
| <u>ESRD due to HTN</u> (n=20 individuals): | 15         | 5          | 0          |
| Caucasian men                              | 75%        | 25%        | 0%         |

PCR and sequencing were conducted as in Example 1. The primers were the same as in Example 2. As shown above, the frequency of the G allele is roughly two times lower (13%) in white men with ESRD due to hypertension compared to a control sample of white men (24%). The frequency of the G allele among white men with hypertension, but no renal failure, 30%, is roughly the same as the control. The genotype frequencies tell a similar story in that the C/C genotype appears to be associated with ESRD due to hypertension, whereas the genotype frequencies of

control and hypertensive white men are quite similar.

These data nicely satisfy the Hardy-Weinberg equilibrium for the control sample. A frequency of 0.76 for the C allele ("p") and 0.24 for the G allele ("q") among control individuals predicts genotype frequencies of 58% C/C, 36% C/G, and 6% G/G at Hardy-Weinberg equilibrium ( $p^2 + 2pq + q^2 = 1$ ). The observed genotype frequencies were 55% C/C, 41% C/G, and 3% G/G, in close agreement with those predicted for Hardy-Weinberg equilibrium.

ESRD, but not essential hypertension, diverges greatly from Hardy-Weinberg equilibrium, consistent with the hypothesis that this SNP is associated with ESRD due to hypertension, but not with essential hypertension itself.

The C845->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene by disrupting the binding site for a number of potential transcriptional regulators whose core recognition sequence consists of the sequence TATC, as follows:

a. The substitution disrupts a GATA\_C (GATA binding site) whose 3' end is at nucleotide #836 on the (-) strand. The binding site consists of the complementary sequence to 5'-NNKNCTTATCN-3' (SEQ ID NO: 6). The C845->G SNP replaces the indicated **C** in the core recognition sequence with a G. Since GATA\_C is a transcriptional activator, the C845->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845->G SNP is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1 pathway. The GATA\_C binding sequence occurs relatively frequently in the genome, 2.62 times per 1000 base pairs in vertebrate genomic DNA.

b. The substitution also results in disruption of a GATA1\_02 (GATA-binding factor 1) binding site whose 3' end is at nucleotide #837 on the (-) strand. The binding site consists of the complementary sequence to 5'-NNCMNTATCNNNN-3' (SEQ ID NO: 7). The C845->G SNP replaces the indicated **C** in the core recognition sequence with a G. Since GATA1\_02 is a

transcriptional activator, the C845->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845->G SNP 5 is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1 pathway. The GATA1\_02 binding sequence occurs relatively frequently in the genome, 2.27 times per 1000 base pairs in vertebrate genomic DNA.

c. There is also disruption of a GATA1\_03 (GATA-binding factor 1) 10 binding site whose 3' end is at nucleotide #837 on the (-) strand. The binding site consists of the complementary sequence to 5'-NCNNTTATCNNNNN-3' (SEQ ID NO: 8). The C845->G SNP replaces the indicated C in the core recognition sequence with a G. Since GATA1\_03 is a transcriptional activator, the C845->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the 15 rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845->G SNP is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1 pathway. The GATA1\_03 binding 20 sequence occurs relatively frequently in the genome, 2.08 times per 1000 base pairs in vertebrate genomic DNA.

d. The substitution results in disruption of a GATA1\_04 (GATA-binding factor 1) binding site whose 3' end is at nucleotide #837 on the (-) strand. The binding site consists of the complementary sequence to 5'- 25 NNNNYTATCWGNN-3' (SEQ ID NO: 9). The C845->G SNP replaces the indicated C in the core recognition sequence with a G. Since GATA1\_04 is a transcriptional activator, the C845->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of 30 this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845->G SNP is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1

pathway. The GATA1\_04 binding sequence occurs relatively frequently in the genome, 1.82 times per 1000 base pairs in vertebrate genomic DNA..

e. In addition, there is a disruption of a GATA2\_02 (GATA-binding factor 2) binding site whose 3' end is at nucleotide #839 on the (-) strand. The 5 binding site consists of the complementary sequence to 5'-TSTTATCWNN-3' (SEQ ID NO: 10). The C845-->G SNP replaces the indicated **C** in the core recognition sequence with a G. This sequence disagrees at only one nucleotide (A841 should be a T) from the ideal, consensus binding site sequence for GATA2\_02, suggesting that it may be functional. Since GATA2\_02 is a transcriptional activator, the C845-->G 10 SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845-->G SNP is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease 15 due to excess signaling through the TGF $\beta$ 1 pathway. It is not known how frequently the GATA2\_02 binding sequence occurs in the genome.

f. There is also disruption of a GATA2\_03 (GATA-binding factor 2) binding site whose 3' end is at nucleotide #839 on the (-) strand. The binding site consists of the complementary sequence to 5'-TNTTAT~~C~~TSN-3' (SEQ ID NO: 11). The 20 C845-->G SNP replaces the indicated **C** in the core recognition sequence with a G. This sequence disagrees at two nucleotides (A841 should be a T; T847 should be a C or G) from the ideal, consensus binding site sequence for GATA2\_03, suggesting that it may be functional. Since GATA2\_03 is a transcriptional activator, the C845-->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII 25 gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845-->G SNP is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1 pathway. It is 30 not known how frequently the GATA2\_03 binding sequence occurs in the genome.

g. In addition there is a disruption of a GATA3\_02 (GATA-binding factor 3) binding site whose 3' end is at nucleotide #839 on the (-) strand. The

binding site consists of the complementary sequence to 5'-TNTTATCTCN-3' (SEQ ID NO: 12). The C845-->G SNP replaces the indicated **C** in the core recognition sequence with a G. This sequence disagrees at two nucleotides (A841 should be a T; T847 should be a C) from the ideal, consensus binding site sequence for 5 GATA3\_02, suggesting that it may be functional. Since GATA3\_02 is a transcriptional activator, the C845-->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845-->G SNP 10 is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1 pathway. It is not known how frequently the GATA3\_02 binding sequence occurs in the genome.

h. The substitution also results in disruption of a GATA3\_03 (GATA-binding factor 3) binding site whose 3' end is at nucleotide #839 on the (-) strand. The binding site consists of the complementary sequence to 5'-TWWKATCTNT-3' (SEQ ID NO: 13). The C845-->G SNP replaces the indicated **C** in the core recognition sequence with a G. This sequence disagrees at only one nucleotide (C840 should be an A or a T) from the ideal, consensus binding site sequence for 20 GATA3\_03, suggesting that it may be functional. Since GATA3\_03 is a transcriptional activator, the C845-->G SNP is predicted to decrease the rate of transcription of the TGF $\beta$ -RII gene. If the rate of transcription of TGF $\beta$ -RII is correlated with the amount of gene product expressed by cells, and if the amount of this receptor affects signaling through the TGF $\beta$ 1 pathway, then the C845-->G SNP 25 is predicted to decrease signaling through the TGF $\beta$ 1 pathway. In other words, this SNP should be protective against disease due to excess signaling through the TGF $\beta$ 1 pathway. It is not known how frequently the GATA3\_03 binding sequence occurs in the genome.

These data suggest that the reference sequence C845 allele contributes 30 significantly towards ESRD as a complication of hypertension. Put differently, the G allele, i.e. the single nucleotide polymorphism at this position, appears to be strongly and specifically protective against ESRD as a complication of hypertension.

**Example 5**G to C Substitution at Position 876 of Human TGF $\beta$ -RII Promoter

Table 7

| ALLELLE FREQUENCIES                 |          |          |
|-------------------------------------|----------|----------|
|                                     | <u>G</u> | <u>C</u> |
| <u>CONTROL</u> (n=58 chromosomes):  | 36       | 22       |
| Caucasian men                       | 62%      | 38%      |
| <u>DISEASE</u>                      |          |          |
| ESRD due to HTN (n=40 chromosomes): | 22       | 18       |
| Caucasian men                       | 55%      | 45%      |

5

Table 8

| GENOTYPE FREQUENCIES                |            |            |            |
|-------------------------------------|------------|------------|------------|
|                                     | <u>G/G</u> | <u>G/C</u> | <u>C/C</u> |
| <u>CONTROL</u> (n=29 individuals):  | 10         | 16         | 3          |
| Caucasian men                       | 34%        | 55%        | 10%        |
| <u>DISEASE</u>                      |            |            |            |
| ESRD due to HTN (n=20 individuals): | 4          | 14         | 2          |
| Caucasian men                       | 20%        | 70%        | 10%        |

PCR and sequencing were conducted as in Example 1. The primers were the same as in Example 2. As demonstrated above, the frequency of the C allele is somewhat higher (45%) in white men with ESRD due to hypertension compared to a control sample of white men (38%).

These data nicely satisfy Hardy-Weinberg equilibrium for the control sample. A frequency of 0.62 for the G allele ("p") and 0.38 for the C allele ("q") among control individuals predicts genotype frequencies of 38% G/G, 47% G/C, and 14% C/C at Hardy-Weinberg equilibrium ( $p^2 + 2pq + q^2 = 1$ ). The observed genotype frequencies were 34% G/G, 55% G/C, and 10% C/C, in close agreement with those predicted for Hardy-Weinberg equilibrium.

ESRD diverges from Hardy-Weinberg equilibrium, with an excess of G/C heterozygotes and a deficiency of G/G homozygotes. These data suggest that the "C" allele contributes moderately towards ESRD. Put differently, the G allele, i.e. the reference allele at this position, appears to be protective against ESRD as a 5 complication of hypertension.

The G876->C SNP is predicted to disrupt a single known transcriptional regulatory site, that of HFH1\_01 (human Forkhead homolog 1; forkhead domain factor HFH-1). The HFH1\_01 consensus binding site sequence consists of the following sequence beginning at nucleotide #872 on the (+) strand: 5-  
10 NAWT**G**TTTATWT-3' (SEQ ID NO: 14). The G876->C SNP replaces the indicated **G** with a C. HFH1\_01 binding sites occur rather rarely, 0.12 times per 1000 base pairs of random genomic sequence in vertebrates, suggesting that this putative transcriptional regulatory site may be functional.

HFH-1 can activate or repress transcription. Consideration of the model for 15 renal failure, namely increased TGF- $\beta$ 1 signaling, would suggest that HFH-1 represses transcription of TGF $\beta$ -RII. The G876->C SNP would therefore be expected to reduce binding affinity of HFH-1 for this site, and thereby relieve repression of the TGF $\beta$ -RII gene.

The G876->C SNP appears to be associated with ESRD due to 20 hypertension, presumably by disrupting a binding site for HFH-1 which in this case would be acting as a transcriptional repressor.

**Example 6**G to T Substitution at Position 945 of Human TGF $\beta$ -RII Promoter

Table 9

| ALLELLE FREQUENCIES                 |     |     |
|-------------------------------------|-----|-----|
|                                     | G   | T   |
| <u>CONTROL</u> (n=52 chromosomes):  | 45  | 7   |
| Caucasian men                       | 87% | 13% |
| <u>DISEASE</u>                      |     |     |
| HYPERTENSION (n=52 chromosomes):    | 45  | 7   |
| Caucasian men                       | 87% | 13% |
| ESRD due to HTN (n=46 chromosomes): | 33  | 13  |
| Caucasian men                       | 72% | 28% |

5

Table 10

| GENOTYPE FREQUENCIES                |     |     |     |
|-------------------------------------|-----|-----|-----|
|                                     | G/G | G/T | T/T |
| <u>CONTROL</u> (n=26 individuals):  | 19  | 7   | 0   |
| Caucasian men                       | 73% | 27% | 0%  |
| <u>DISEASE</u>                      |     |     |     |
| HYPERTENSION (n=26 individuals):    | 19  | 7   | 0   |
| Caucasian men                       | 73% | 27% | 0%  |
| ESRD due to HTN (n=23 individuals): | 11  | 11  | 1   |
| Caucasian men                       | 48% | 48% | 4%  |

PCR and sequencing were conducted as in Example 1. The sense primer was 5'-GGACATATCTGAAAGAGAAAGGGGG-3' (SEQ ID NO: 15) and the antisense primer was 5'-TTGGGAGTCACCTGAATGCTTG-3' (SEQ ID NO: 16). As demonstrated above, the frequency of the T allele is over twice as high among white men with ESRD due to hypertension (28%) compared to a control sample of

white men (13%). The allele and genotype frequencies are the same for the control sample and for white men with essential hypertension but no renal failure, suggesting that the T allele is specific for ESRD.

These data satisfy Hardy-Weinberg equilibrium for the control sample and 5 white men with hypertension. A frequency of 0.87 for the G allele ("p") and 0.13 for the T allele ("q") among control individuals predicts genotype frequencies of 76% G/G, 22% G/T, and 2% T/T at Hardy-Weinberg equilibrium ( $p^2 + 2pq + q^2 = 1$ ). The observed genotype frequencies were 73% G/G, 27% G/T, and 0% T/T, in reasonably close agreement with those predicted for Hardy-Weinberg equilibrium.

10 ESRD diverges from Hardy-Weinberg equilibrium, with an excess of G/T heterozygotes and a deficiency of G/G homozygotes. These data suggest that the "T" allele contributes strongly and specifically towards ESRD. Put differently, the G allele, i.e. the reference allele at this position, appears to be protective against ESRD as a complication of hypertension.

15 The G945->T SNP does not disrupt any known transcriptional regulatory site. To be consistent with the model of increased TGF $\beta$ 1 signaling as a cause of renal failure, it is expected that an as yet unknown transcriptional repressor binds to this region of the TGF $\beta$ -RII promoter.

20 The G945->T SNP appears to be associated specifically with ESRD due to hypertension in white men. It is hypothesized that this SNP disrupts the binding site for an as yet undescribed transcriptional repressor of the TGF $\beta$ -RII gene.

## Example 7

G to W (A or T) Substitution at Position 983 of Human TGF $\beta$ -RII Promoter

Table 11

| ALLELLE FREQUENCIES                 |          |          |          |
|-------------------------------------|----------|----------|----------|
|                                     | <u>G</u> | <u>A</u> | <u>T</u> |
| <u>CONTROL</u> (n=52 chromosomes):  | 45       | 7        | 0        |
| Caucasian men                       | 87%      | 13%      | 0%       |
| <u>DISEASE</u>                      |          |          |          |
| HYPERTENSION (n=54 chromosomes):    | 50       | 0        | 4        |
| Caucasian men                       | 93%      | 0%       | 7%       |
| ESRD due to HTN (n=46 chromosomes): | 40       | 0        | 6        |
| Caucasian men                       | 87%      | 0%       | 13%      |

5 Table 12

| GENOTYPE FREQUENCIES                |            |            |            |
|-------------------------------------|------------|------------|------------|
|                                     | <u>G/G</u> | <u>G/A</u> | <u>G/T</u> |
| <u>CONTROL</u> (n=26 individuals):  | 19         | 7          | 0          |
| Caucasian men                       | 73%        | 27%        | 0%         |
| <u>DISEASE</u>                      |            |            |            |
| HYPERTENSION (n=27 individuals):    | 23         | 0          | 4          |
| Caucasian men                       | 85%        | 0%         | 15%        |
| ESRD due to HTN (n=23 individuals): | 17         | 0          | 6          |
| Caucasian men                       | 74%        | 0%         | 26%        |

PCR and sequencing were conducted as in Example 1. The primers were the same as in Example 6. Most SNPs are biallelic, but the G983-->W SNP is unusual in being triallelic. The frequency of the reference allele, G, is the same for the 10 control and both disease categories: 87% in white male controls, compared to 93% in white men with hypertension, and 87% in white men with ESRD due to hypertension. The A allele, present at low frequency in the control population (13%), does not figure at all in either hypertension or ESRD due to hypertension.

Instead, the T allele appears in the sample with hypertension (7%), and is nearly twice as high among patients with ESRD due to hypertension (13%).

The most straightforward interpretation of these results is that the T allele contributes directly to hypertension, as well as to its complication, ESRD. The G and the A alleles appear to be protective against hypertension as well as ESRD due to hypertension.

The control sample approximates Hardy-Weinberg equilibrium. A frequency of 0.84 for the G allele ("p") and 0.13 for the A allele ("q") among control individuals predicts genotype frequencies of 76% G/G, 22% G/A, and 2% A/A at Hardy-Weinberg equilibrium ( $p^2 + 2pq + q^2 = 1$ ). The observed genotype frequencies were 73% G/G, 27% G/A, and 0% A/A, in very close agreement with those predicted for Hardy-Weinberg equilibrium.

The two disease categories diverge greatly from Hardy-Weinberg equilibrium, since they possess the T allele which does not appear in the control sample at all. These data strongly suggest that the T allele is associated with hypertension, as well as ESRD due to hypertension.

The G983-->W SNP is predicted to disrupt a potential RFX1\_02 (X-box binding protein RFX1) binding site whose 3' terminus ends at nucleotide 972 on the (-) strand. The consensus RFX1\_02 binding site consists of the sequence complementary to 5'-NNGTTRCYNNNN~~G~~YNACNN-3' (SEQ ID NO: 17). Both the G983-->T and G983-->A forms of this triallelic SNP replace the indicated **G** in the core recognition sequence. Why the T allele should be associated with disease but not the A allele is unclear. RFX1\_02 binding sites occur somewhat frequently, 0.95 matches per 1000 base pairs of random genomic sequence in vertebrates.

The G983-->W SNP is complex in that it is triallelic. Only the T allele appears to be associated with hypertension, as well as ESRD due to hypertension. Why the A allele should be protective is unclear. The only known transcriptional regulatory site affected by this polymorphism is an RFX1\_02 binding site. To be consistent with the model that progression of chronic renal failure involves increased TGF- $\beta$ 1 signaling, RFX1\_02 would be expected to function as a transcriptional repressor at this position. However, the association of the T allele with hypertension is unexpected and suggests a novel mechanism for hypertension involving signaling

through the type II TGF- $\beta$ 1 receptor.

## CONCLUSION

In light of the detailed description of the invention and the examples presented above, it can be appreciated that the several aspects of the invention are 5 achieved.

It is to be understood that the present invention has been described in detail by way of illustration and example in order to acquaint others skilled in the art with the invention, its principles, and its practical application. Particular formulations and processes of the present invention are not limited to the descriptions of the 10 specific embodiments presented, but rather the descriptions and examples should be viewed in terms of the claims that follow and their equivalents. While some of the examples and descriptions above include some conclusions about the way the invention may function, the inventor does not intend to be bound by those conclusions and functions, but puts them forth only as possible explanations.

15 It is to be further understood that the specific embodiments of the present invention as set forth are not intended as being exhaustive or limiting of the invention, and that many alternatives, modifications, and variations will be apparent to those of ordinary skill in the art in light of the foregoing examples and detailed description. Accordingly, this invention is intended to embrace all such alternatives, 20 modifications, and variations that fall within the spirit and scope of the following claims.

Table 13

| Gene             | Region   | Location | Wild Type | Variant | SEQ ID |
|------------------|----------|----------|-----------|---------|--------|
| TGF $\beta$ -RII | Promoter | 796      | A         | C       | 1      |
|                  |          | 820      | A         | C       | 1      |
|                  |          | 845      | C         | G       | 1      |
|                  |          | 876      | G         | C       | 1      |
|                  |          | 945      | G         | T       | 1      |
|                  |          | 983      | G         | W       | 1      |

**Table 14**

| Gene             | Region   | Location | Wild Type | Variant | SEQ ID |
|------------------|----------|----------|-----------|---------|--------|
| TGF $\beta$ -RII | Promoter | 796      | A         | C       | 1      |
|                  |          | 983      | G         | W       | 1      |

What is claimed is:

1. A method for diagnosing a genetic susceptibility for a disease, condition, or disorder in a subject comprising:  
obtaining a biological sample containing nucleic acid from said subject; and  
analyzing said nucleic acid to detect the presence or absence of a single  
nucleotide polymorphism in the TGF $\beta$ -RII gene, wherein said single nucleotide  
polymorphism is associated with a genetic predisposition for a disease selected  
from the group consisting of hypertension and end-stage renal disease due to  
hypertension.
2. The method of claim 1, wherein the gene TGF $\beta$ -RII comprises SEQ ID NO: 1.
3. The method of claim 1, wherein said nucleic acid is DNA, RNA, cDNA or  
mRNA.
4. The method of claim 2, wherein said single nucleotide polymorphism is located  
at position 796, 820, 845, 876, 945 or 983 of SEQ ID NO: 1.
5. The method of claim 4, wherein said single nucleotide polymorphism is a  
selected from the group consisting of A820-> C, T820->G, C845->G, G845->C,  
G876->C, C876->G, G945->T, C945->A, G983->A, G983->T, C983->A, and  
C983->T.
6. The method of claim 1, wherein said analysis is accomplished by sequencing,  
mini sequencing, hybridization, restriction fragment analysis, oligonucleotide  
ligation assay or allele specific PCR.

7. An isolated polynucleotide comprising at least 10 contiguous nucleotides of SEQ ID NO: 1, or the complements thereof, and containing at least one single nucleotide polymorphism at position 796, 820, 845, 876, 945 or 983 of SEQ ID NO: 1 wherein said at least one single nucleotide polymorphism is associated with a disease selected from the group consisting of hypertension and end stage renal disease due to hypertension.  
5
8. The isolated polynucleotide of claim 7, wherein at least one single nucleotide polymorphism is selected from the group consisting of A820-> C, T820->G, C845->G, G845->C, G876->C, C876->G, G945->T, C945->A, G983->A, G983->T, C983->A, and C983->T.
9. The isolated polynucleotide of claim 7, wherein said at least one single nucleotide polymorphism is located at the 3= end of said nucleic acid sequence.
10. The isolated polynucleotide of claim 7, further comprising a detectable label.
11. The isolated nucleic acid sequence of claim 10, wherein said detectable label is selected from the group consisting of radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.
12. A kit comprising at least one isolated polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO: 1 or the complement thereof, and containing at least one single nucleotide polymorphism associated with a disease, condition, or disorder selected from the group consisting of hypertension and end stage renal disease due to hypertension; and instructions for using said polynucleotide for detecting the presence or absence of said at least one single nucleotide polymorphism in said nucleic acid.  
5
13. The kit of claim 12 wherein said at least one single nucleotide polymorphism is located at position 796, 820, 845, 876, 945 or 983 of SEQ ID NO: 1.

14. The kit of claim 13 wherein said at least one single nucleotide polymorphism is selected from the group consisting of A820-> C, T820->G, C845->G, G845->C, G876->C, C876->G, G945->T, C945->A, G983->A, G983->T, C983->A, and C983->T.
15. The kit of claim 12, wherein said single nucleotide polymorphism is located at the 3= end of said polynucleotide.
16. The kit of claim 12, wherein said polynucleotide further comprises at least one detectable label.
17. The kit of claim 16, wherein said label is chosen from the group consisting of radionuclides, fluorophores or fluorochromes, peptides enzymes, antigens, antibodies, vitamins or steroids.
18. A kit comprising at least one polynucleotide of at least 10 contiguous nucleotides of SEQ ID NO: 1 or the complement thereof, wherein the 3= end of said polynucleotide is immediately 5= to a single nucleotide polymorphism site associated with a genetic predisposition to disease, condition, or disorder selected from the group consisting of hypertension and end stage renal disease due to hypertension; and instructions for using said polynucleotide for detecting the presence or absence of said single nucleotide polymorphism in a biological sample containing nucleic acid.
19. The kit of claim 18, wherein said at least one polynucleotide further comprises a detectable label.
20. The kit of claim 19, wherein said detectable label is chosen from the group consisting of radionuclides, fluorophores or fluorochromes, peptides, enzymes, antigens, antibodies, vitamins or steroids.
21. A method for treatment or prophylaxis in a subject comprising:

obtaining a sample of biological material containing nucleic acid from a subject; analyzing said nucleic acid to detect the presence or absence of at least one single nucleotide polymorphism in SEQ ID NO: 1 or the complement thereof

5 associated with a disease, condition, or disorder selected from the group consisting of hypertension and end stage renal disease due to hypertension; and treating said subject for said disease, condition or disorder.

22. The method of claim 21 wherein said nucleic acid is selected from the group consisting of DNA, cDNA, RNA and mRNA.

23. The method of claim 21, wherein said at least one single nucleotide polymorphism is located at position 796, 820, 845, 876, 945 or 983 of SEQ ID NO: 1.

24. The method of claim 21 wherein said at least one single nucleotide polymorphism is selected from the group consisting of A820-> C, T820->G, C845->G, G845->C, G876->C, C876->G, G945->T, C945->A, G983->A, G983->T, C983->A, and C983->T.

25. The method of claim 21 wherein said treatment counteracts the effect of said at least one single nucleotide polymorphism detected.

## SEQUENCE LISTING

<110> DzGenes LLC

<120> DIAGNOSTIC POLYMORPHISMS OF TGF-beta-RII PROMOTOR

<130> DZG 2177.1

<150> US 60/191,737

<151> 2000-03-24

<160> 17

<170> PatentIn version 3.0

<210> 1

<211> 1883

<212> DNA

<213> Homo sapiens

<220>

<221> gene

<222> (1)..(1883)

<223> TGF-beta RII

<220>

<221> promoter

<222> (1)..(1883)

<223> TGF-beta RII Promotor

|   |  |     |
|---|--|-----|
| <400> 1   |  |     |
| cccatcaaag aagttatgat tcaatccacg aagaccagga gttggcgaaa tgaagaaaaa   |  | 60  |
| aaggtaagtc gaaggaagtc ctctctgggg aaggctctaa gcataaaggg caggaggatt   |  | 120 |
| acagaggcat atctcgaaat ttggagaagg ctttcagtaa gcaaggagaa gccaaatgaa   |  | 180 |
| agtttacgga gagttggagg cttgaagaca ccgttcaagg atctggtttt tatcttctct   |  | 240 |
| ttattctcaa gagcttagtg ggaagccatt aaatgatttt aatcaaggag gggttggta    |  | 300 |
| taaactagtt ttgttaattt tgaaaaatct gaattcactc tcgtttgaga aactgagtga   |  | 360 |
| aagagcccag aacggccgtg ctgaggggtga ctccctggaa gactccttaa ccacaagcca  |  | 420 |
| tggcagtggc atgggctggt ggcagaagag ggaataggga gaagatttgg aactcaatct   |  | 480 |
| tcctccattg acaaagtac tccagctttg gcaaggcaat taattgggtgg gaaagaagat   |  | 540 |
| gcctagccct cctgatttca ctgcacttcc tgcatcttca acatgagtac tgggaagtgg   |  | 600 |
| caaaaacaatc cagagggcagg cttgggtgct aggtggagca tgagttaaaa ttccaggatg |  | 660 |
| aagcaaatga acacttagaa tgacaggaaa gatttggag ttgggtttgg gggagggcta    |  | 720 |
| tttaccttta ttccctggag accctggcac aaaccctgcc tctgcaatct tcctctcagg   |  | 780 |
| taaaggaatt cattaaatga attgctagaa gatctactga ccagagggct gtacagaatc   |  | 840 |

|   |      |
|---|------|
| atatcttga gagtggaag tagttgatc acatagttt aatccaaatc aggacatatc       | 900  |
| tgaaagagaa agggggttct attaatattt aaactacaaa acatgtacac caggaatgtc   | 960  |
| ttggccaaat ctgggtgccc tagcaagaaa gaaatttga aagtttatgc tttctgctc     | 1020 |
| ccatgttacc ccgttgcac atgagagggt aagtattctc tttcttacc tgcattaagg     | 1080 |
| gaataaaaagc acaagcattc aggtgactcc caacccactt ttaattttac agtttctgct  | 1140 |
| atactctata cattctgaaa attacatttc ccaccactat acttcgtat aggtgatcat    | 1200 |
| ttacaattac tcactgactc agtcccggga agaggcggtg caaaatggac gctctatcca   | 1260 |
| ggtgctcatt agaaatgcag aatctctgcc tgcctcctag acctactgaa tttagaatctg  | 1320 |
| cattttaaa taagatttcc aggtgatcaa tatgtacatt aaaacttgag aaaaacctct    | 1380 |
| agacttcgac ctaaaagaaaa acattttaca acttgacagt gtatgcacat acatacatgc  | 1440 |
| atataagacac aactgaagca caaatttaat gaagttagat ttaccgttac tattttattt  | 1500 |
| ggaaagaaat gtgctcgca ctcaatagat tggagtattc actcctggat ctcaacttgc    | 1560 |
| aatttgaaaa cgcatctcta aagcacctag gagcaatctg aagaaagctg aggggaggcg   | 1620 |
| gcagatgttc tgcatttacta gggaaaacgt ggacgtttc tgggttact ttgtgaactg    | 1680 |
| tgtgcactta gtcatttttgc agtaaatact tggagcgagg aactcctgag tgggtgtggga | 1740 |
| gggcgggtgag gggcagctga aagtcggcca aagctctcg aagggctgggt ctaggaaaca  | 1800 |
| tgattggcag ctacgagaga gctaggggct ggacgtcgag gagagggaga aggctctcg    | 1860 |
| gcggagagag gtcctgcccc gct   | 1883 |

<210> 2  
 <211> 19  
 <212> DNA  
 <213> Artificial

<220>  
 <221> misc\_feature  
 <222> (1)..(19)  
 <223> Primer

<400> 2  
 ggagtgggt ttggggag 19

<210> 3  
 <211> 22  
 <212> DNA  
 <213> Artificial

<220>  
 <221> misc\_feature  
 <222> (1)..(22)  
 <223> Primer

<400> 3  
tcttgcttagg gcaaccagat tg

22

<210> 4  
<211> 13  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(13)

<220>  
<221> variation  
<222> (8)..(8)

<220>  
<221> variation  
<222> (1)..(13)  
<223> n=a, c, g or t

<400> 4  
gtcatnnwnnn nnn

13

<210> 5  
<211> 13  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(13)

<220>  
<221> variation  
<222> (8)..(8)

<220>  
<221> variation  
<222> (1)..(13)  
<223> n=a, c, g or t

<400> 5  
gtcatnnwnnn nnn

13

<210> 6  
<211> 11  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(11)

<220>  
<221> variation  
<222> (10)..(10)

<220>  
<221> variation  
<222> (1)..(11)  
<223> n=a, c, g or t

<400> 6  
nnkncttatac n

11

<210> 7  
<211> 14  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(14)

<220>  
<221> variation  
<222> (9)..(9)

<220>  
<221> variation  
<222> (1)..(14)  
<223> n=a, c, g or t

<400> 7  
nncmntatcn nnnn

14

<210> 8  
<211> 14  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(14)

<220>  
<221> variation  
<222> (9)..(9)

<220>  
<221> variation  
<222> (1)..(14)  
<223> n=a, c, g or t

<400> 8  
ncnnttatacn nnnn

14

<210> 9

<211> 13  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(13)

<220>  
<221> variation  
<222> (9)..(9)

<220>  
<221> variation  
<222> (1)..(13)  
<223> n=a, c, g or t

<400> 9  
nnnnytatcw gnn

13

<210> 10  
<211> 10  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(10)

<220>  
<221> variation  
<222> (7)..(7)

<220>  
<221> variation  
<222> (9)..(10)  
<223> n=a, c, g or t

<400> 10  
tsttatacwnn

10

<210> 11  
<211> 10  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(10)

<220>  
<221> variation  
<222> (7)..(7)

<220>  
<221> variation  
<222> (1)..(10)

<223> n=a, c, g or t

<400> 11  
tnttatctsn

10

<210> 12  
<211> 10  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(10)

<220>  
<221> variation  
<222> (7)..(7)

<220>  
<221> variation  
<222> (1)..(10)  
<223> n=a, c, g or t

<400> 12  
tnttatctcn

10

<210> 13  
<211> 10  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(10)

<220>  
<221> variation  
<222> (9)..(9)  
<223> n=a, c, g or t

<220>  
<221> variation  
<222> (7)..(7)

<400> 13  
twwkatctnt

10

<210> 14  
<211> 12  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(12)

<220>  
<221> variation  
<222> (5)..(5)

<220>  
<221> variation  
<222> (1)..(1)  
<223> n=a, c, g or t

<400> 14  
nawtgtttat wt

12

<210> 15  
<211> 25  
<212> DNA  
<213> Artificial

<220>  
<221> misc\_feature  
<222> (1)..(25)  
<223> Primer

<400> 15  
ggacatatct gaaagagaaaa gggggg

25

<210> 16  
<211> 22  
<212> DNA  
<213> Artificial

<220>  
<221> misc\_feature  
<222> (1)..(22)  
<223> Primer

<400> 16  
ttgggagtca cctgaatgct tg

22

<210> 17  
<211> 18  
<212> DNA  
<213> Homo sapiens

<220>  
<221> primer\_bind  
<222> (1)..(18)

<220>  
<221> variation  
<222> (12)..(12)

<220>  
<221> variation  
<222> (1)..(18)

<223> n=a, c, g or t

<400> 17  
nngttrcynn ngynacnn

18

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/09583

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68, C12P 19/34 C07H 21/04  
US CL : 435/6, 91.1, 91.2, 536/24.3, 24.33

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.2, 536/24.3, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Continuation Sheet

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|------------|---|-----------------------|
| A          | FREEDMAN et al. Genetic linkage analysis of growth factor loci and end-stage renal disease in African Americans. Kidney International. 1997, Vol. 51, pages 819-825, see whole document.                        | 1-25                  |
| A          | CAMBIEN et al. Polymorphisms of the transforming growth factor-B1 gene in relation to myocardial infarction and blood pressure. Hypertension. November 1996, Vol. 28, No. 5, pages 881-887, see whole document. | 1-25                  |
| A          | BAE et al. Characterization of the promoter region of the human transforming growth factor-B type II receptor gene. J. of Bio. Chem. 08 December 1995, Vol. 270, No. 49, pages 29460-29468, see whole document. | 1-25                  |
| A          | SYRRIS et al. Transforming growth factor-B1 gene polymorphisms and coronary artery disease. Clinical Science. 1998, Vol. 95, pages 659-667, see whole document.   | 1-25                  |

Further documents are listed in the continuation of Box C.

See patent family annex.

|   |     |  |
|---|-----|--|
| Special categories of cited documents:  | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| "A" document defining the general state of the art which is not considered to be of particular relevance  | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| "E" earlier application or patent published on or after the international filing date   | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" | document member of the same patent family  |
| "O" document referring to an oral disclosure, use, exhibition or other means  |     |  |
| "P" document published prior to the international filing date but later than the priority date claimed  |     |  |

Date of the actual completion of the international search

Date of mailing of the international search report

17 JUL 2001

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231  
Facsimile No. (703)305-3230

Authorized officer  
Alexander H. Spiegler  
Telephone No. (703) 308-0196

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US01/09583

**Continuation of B. FIELDS SEARCHED Item3:**  
Medline, Biosis, Embase, Cplus, Biotechds, WEST, Sequence Search

**Search Terms:** TGFB-RII, transforming growth factor beta, receptor, end stage renal disease, esrd, niddm, hypertension, polymorphism, mutation, genotype